Homework Data Visualization by R

gcat332(gust)

2023-12-02

Explore data

```
library(tidyverse)
head(mpg)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

Create 5 charts from mpg data

1. Histogram chart (one variable : continuous (number))

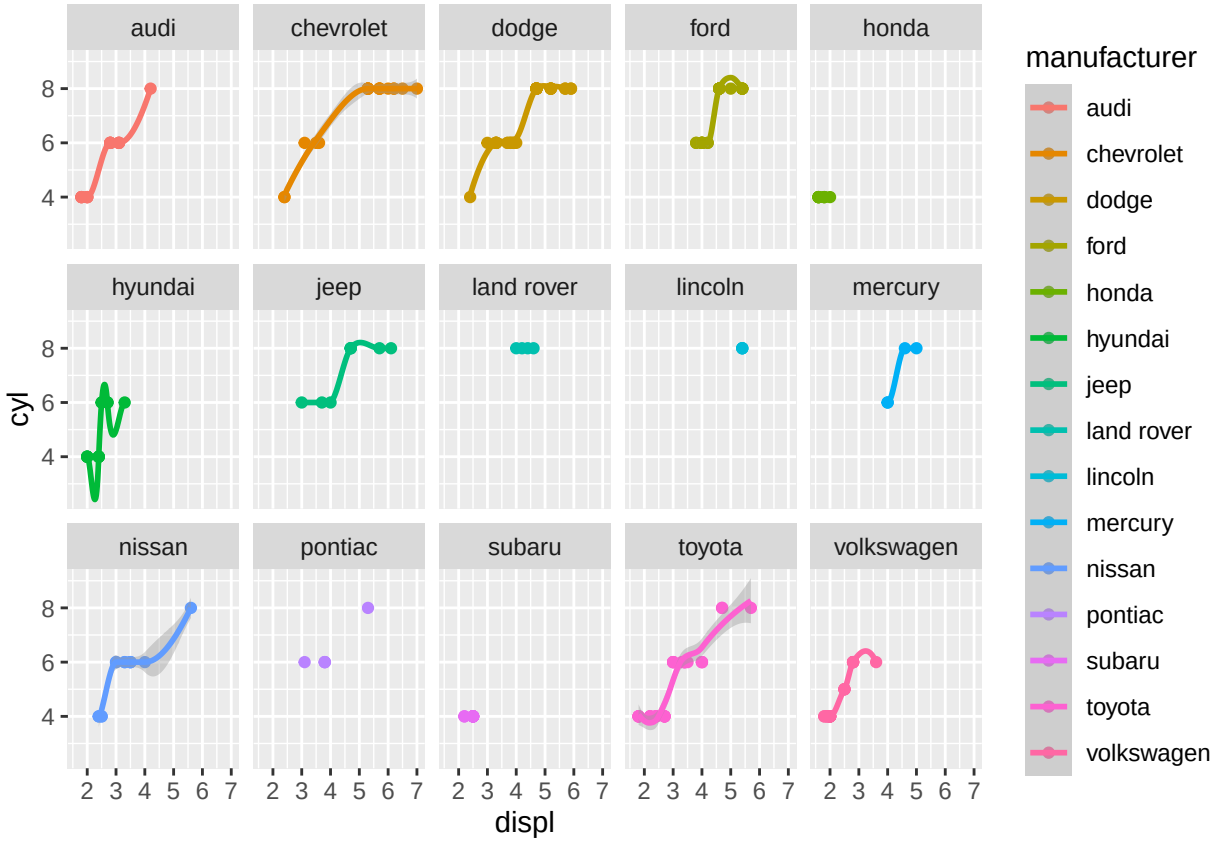
```
## year of manufacture by manufacturer, fill color by model
ggplot(mpg,aes(displ,fill=model)) +
  geom_histogram() +
  facet_wrap(~manufacturer,ncol = 3)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

2. Scatter chart (two variable : continuous (number) x continuous (number))

```
## correlation between engine displacement and number of cylinders group by manufacturer
ggplot(mpg,aes(displ,cyl,color = manufacturer)) +
  geom_point() +
  geom_smooth() +
  facet_wrap(~manufacturer,3)
```

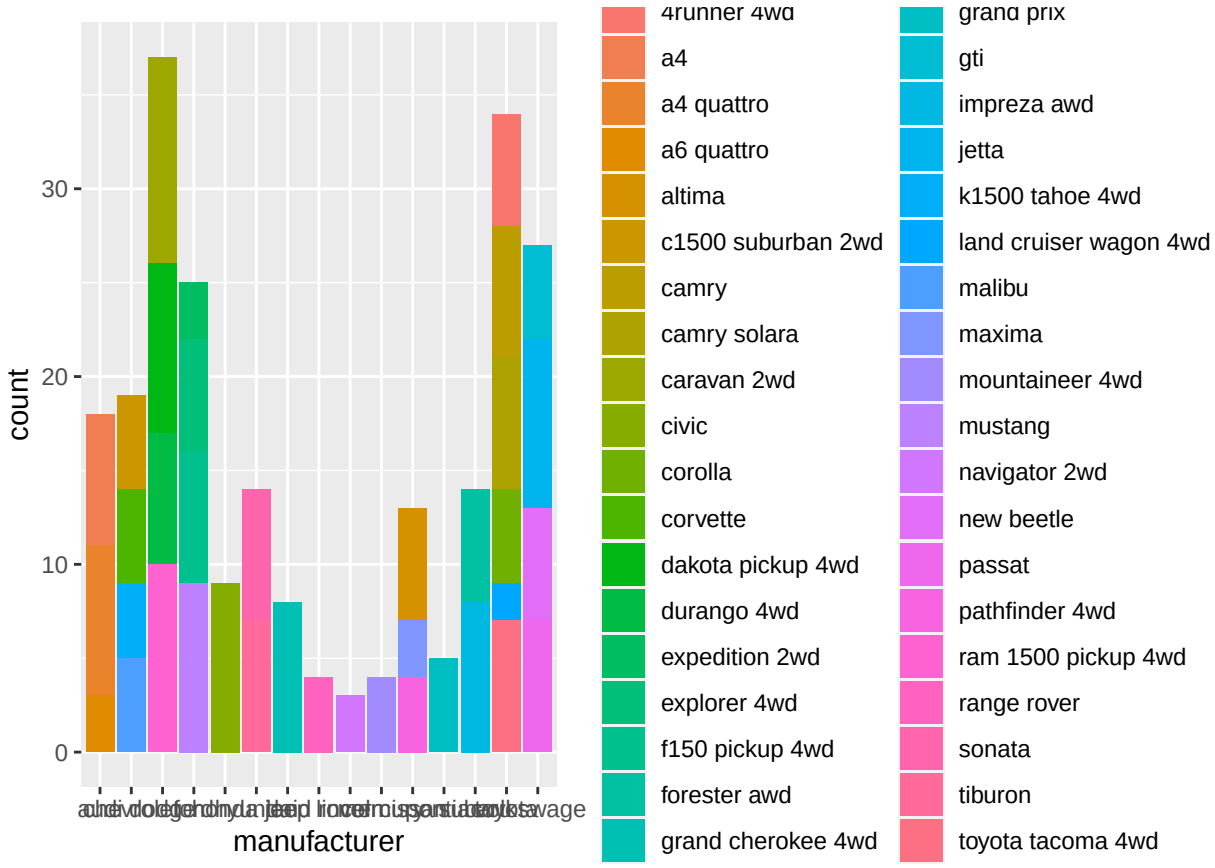
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



3

. Bar chart (one variable : discrete (factor))

```
## Quantity of model per each manufacturer
ggplot(mpg,aes(manufacturer,fill = model)) +
  geom_bar(potision = 'stack')
```

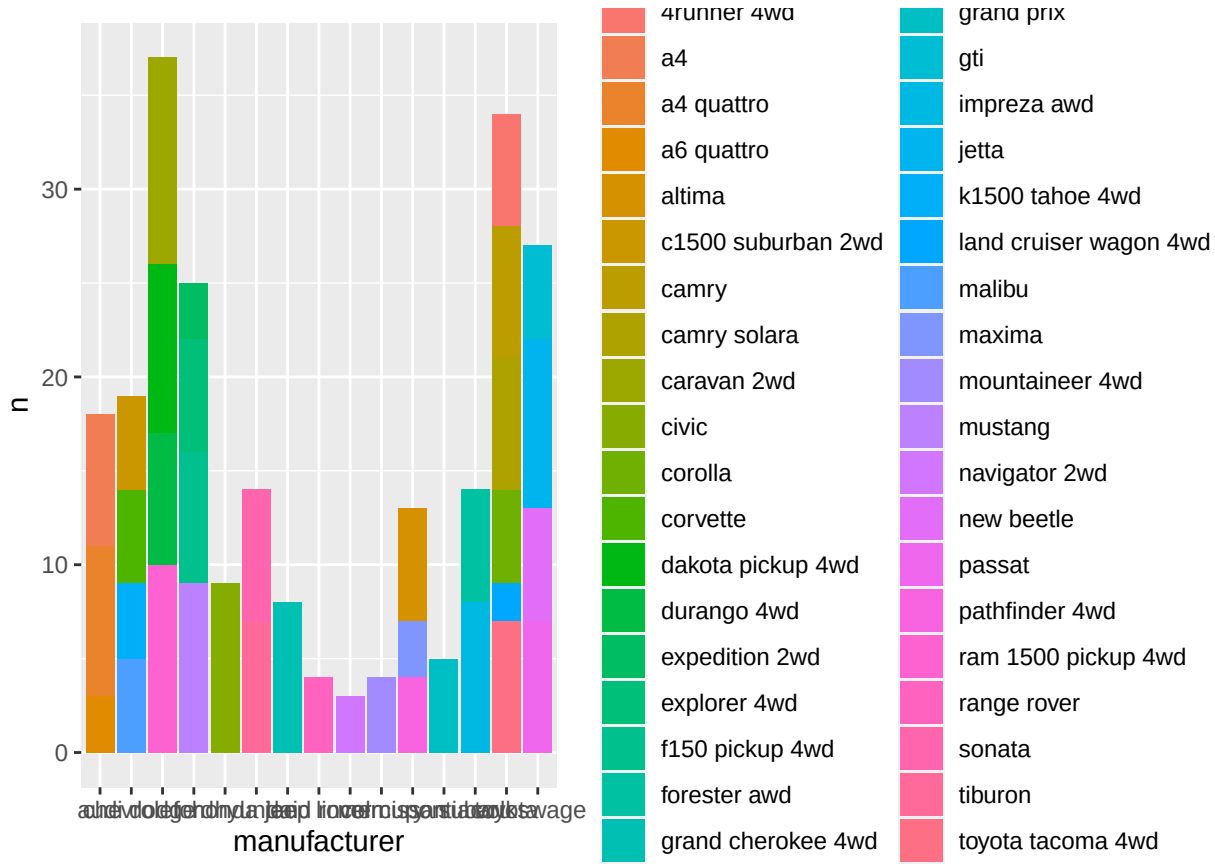


4

. Column chart (two variable : discrete (factor) x continuous (number))

```
## Quantity of model per each manufacturer but with summarise() method
mpg %>%
  group_by(manufacturer,model) %>%
  summarise(n=n())%>%
  ggplot(aes(manufacturer,n,fill=model)) +
  geom_col(potision='stack')
```

```
## `summarise()` has grouped output by 'manufacturer'. You can override using the
## `.groups` argument.
```



5

. Scatter chart with facet 2 factors

```
## find correlation of city miles and highway miles grouping by fuel type and type of transmission
ggplot(mpg,aes(cty,hwy,color = fl)) +
  geom_point() +
  geom_smooth() +
  facet_grid(fl~trans)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```